

Tema 1: Fundamentos de diseño y evaluación de computadores

Departament d'Arquitectura de Computadors

Facultat d'Informàtica de Barcelona

Universitat Politècnica de Catalunya





"I think there's a world market for about 5 computers."

(Thomas J. Watson, Chairman of the Board, IBM, circa 1948)

"In the future, computers may weigh no more than 1.5 tonnes."

– Popular mechanics, 1949

"There is no reason for any individual to have a computer in his home."

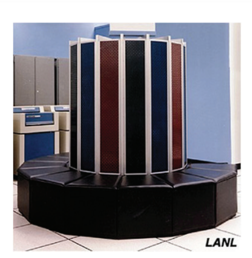
(Ken Olson, President, Digital Equipment Corporation, 1977)

Índice

- Introducción
- Coste
- Rendimiento
- Consumo
- Fiabilidad

Evolución de los computadores

	CRAY 1	Lenovo T61
Año (instalación)	1976	2008
CPU	Custom (circuitos discretos)	Intel Core 2 Duo T7300 (Merom)
Características	Procesador vectorial Sin Cache MP SRAM	2 cores IA32 segmentado L1I (2×32KB) & L1D(2×32KB) + L2 (4MB) MMX, SSE(1,2,3,3S),EM64T,VT-x
Consumo	115.000 W	100 W (CPU 35W)
Dimensiones	∅: 262,89 cm – 143,51 Alt: 195,58 – 48,26	3,0 × 34,0 × 24,0 cm
Peso	5.500 Kg	2,5 Kg
Coste	8,86 millones dólares (IPC acumulado España 910,7%)	1.500€
Memoria	8 MB	2 GB
Disco	2,5 GB (1 millón $, 1976)	120 GB (50€, abril 2010)
Rendimiento	160 MFLOPS (pico), 50 MFLOPS sostenido	10 GFLOPS (linpack)
Frecuencia	80 MHz	2 GHz
	Refrigerado con freón	Portátil



¡32años!



¡32años!



Tipos de computadores

	Sobremesa	Servidor	Supercomputador	Embedded
Coste del sistema	500-5.000€	5.000-5.000.000€	5.000.000-????€	1-100.000€
Precio CPU	50-500€	200-10.000€	200-2.000€	0,01-100€
Puntos críticos en el diseño del sistema	Coste/rendimiento, rendimiento gráficos	Throughput, disponibilidad, escalabilidad	Rendimiento en coma flotante, capacidad almacenamiento	Precio, consumo, rendimiento en aplicaciones especificas
Aplicaciones	Ofimática, Ocio, Estación de trabajo, Desarrollo software, …	Servidor web, Bases de Datos, …	Geofísica, Meteorología, Diseño de aviones, …	Teléfonos móviles, Lectores BluRay, Automóviles, …
#cores	1-8	8-10.000	1.280-294.912 (1)	1-80
Memoria	Gbytes	Tbytes	Pbytes	Mbytes
Disco	Tbytes	Pbytes	Pbytes	Gbytes

(1) Top500 Nov/2009

Pesos y Medidas

- Hay un cierto desconcierto a la hora de utilizar los prefijos de Medida

Nombre	Símbolo	2^x		10^x		error
Kilo	**K**	2^{10}	1024	10^3	1000	2,4%
Mega	**M**	2^{20}	1048576	10^6	1000000	4,9%
Giga	**G**	2^{30}	1073741824	10^9	1000000000	7,4%
Tera	**T**	2^{40}	1099511627776	10^{12}	1000000000000	10,0%
Peta	**P**	2^{50}	1125899906842624	10^{15}	1000000000000000	12,6%
Exa	**E**	2^{60}	1152921504606846976	10^{18}	1000000000000000000	15,3%
Zetta	**Z**	2^{70}	1180591620717411303424	10^{21}	1000000000000000000000	18,1%
Yotta	**Y**	2^{80}	1208925819614629174706176	10^{24}	1000000000000000000000000	20,1%
Xenta/Xora/Bronto		2^{90}	1257282852399214341694423O4	10^{27}	1000000000000000000000000000	25,7%

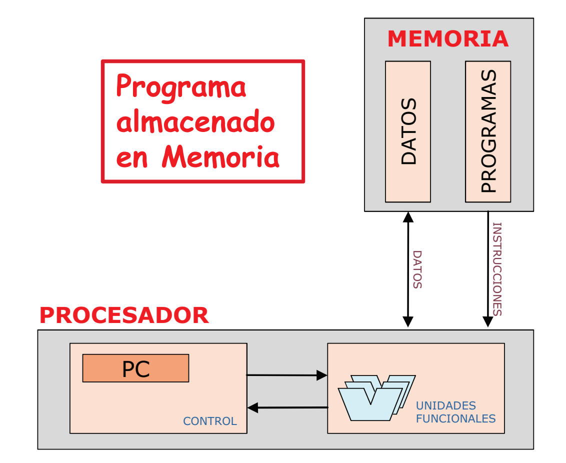
- Para los prefijos binarios existe la **norma ISO/CEI**, aunque no está suficientemente extendida. Por ejemplo, $\mathbf{2^{20}}$ se denomina **Mebi** y usa el símbolo **Mi**.

Pesos y Medidas

Problema: no todo el mundo quiere decir lo mismo cuando utiliza la misma palabra:

- Los Hercios (Hz) se miden en potencias de 10: un procesador a 1 GigaHercio (GHz) va a 1.000.000.000 Hz.
- La velocidad de transmisión se mide en potencias de 10: un MP3 stream a 128 Kb/s transmite 128.000 bits por segundo, una conexión ADSL de 12 Mb/s acepta un máximo de 12.000.000 bits por segundo.
- El ancho de banda de los buses también se mide en potencias de 10
- La Memoria RAM siempre se mide en potencias de 2: 1GB de RAM es 2^{30} bytes de RAM.
- Los discos duros (HD) utilizan potencias de 10.
  - Un HD de 30GB tiene $30 \cdot 10^9$ bytes (aproximadamente $28 \cdot 2^{30}$).
  - No es marketing, sino tradición: la estructura física de los discos (platos, pistas, sectores) no tiene por qué ser potencia de 2.
  - Además, el SO suele indicar el tamaño del disco en potencias de 2.
    - ✓ Por tanto, si compramos un portátil con 1GB de RAM y 30 GB de HD, Windows nos dirá que tiene 1GB de RAM y 28GB de disco duro.

Máquina Von Neumann



John Von Neumann (matemático húngaro) "First Draft of a report on the EDVAC" contract n. W-670-DRD-492 Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, **June 1945**

Sólo aparece un autor aunque el report es resultado de múltiples horas de discusión del grupo que diseño el ENIAC, en el que Von Neumann era sólo un visitante. Prosper Eckert y John Mauchly (diseñadores principales del ENIAC) abandonaron la Moore School por este hecho.

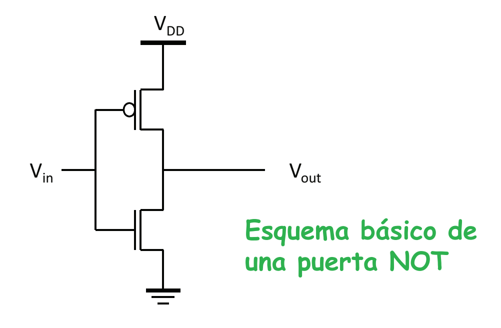
- **Memoria accesible por dirección.**
- **Las posiciones de memoria se pueden leer/escribir las veces que sean necesarias.**
- **Tanto los datos como las instrucciones se almacenan en memoria.**
- **No existe ninguna señal para diferenciar en memoria datos de instrucciones.**
- **Las instrucciones se ejecutan en secuencia.**
- **Existe un registro (PC) que apunta siempre a la instrucción en ejecución.**
- **Existen instrucciones explícitas para romper el secuenciamiento.**
- **Las instrucciones son imperativas, especifican cómo obtener los operandos, qué operación hay que realizar y dónde dejar el resultado.**

Evaluación de un sistema informático

- **Métricas para caracterizar un sistema informático:**
  - **Coste**:
    - ✓ Tamaño del die (dado).
    - ✓ Complejidad: esfuerzo requerido en el diseño, validación y fabricación del procesador.
    - ✓ Coste ambiental y social.
  - **Rendimiento**: Inversa del tiempo que tarda en completarse una tarea.
  - **Consumo**: Energía consumida por unidad de tiempo (watios). Normalmente mayor rendimiento requiere mayor consumo.
  - **Fiabilidad**: Tiempo entre fallos/reparaciones. Sistemas tolerantes a fallos.

Tecnología de Fabricación

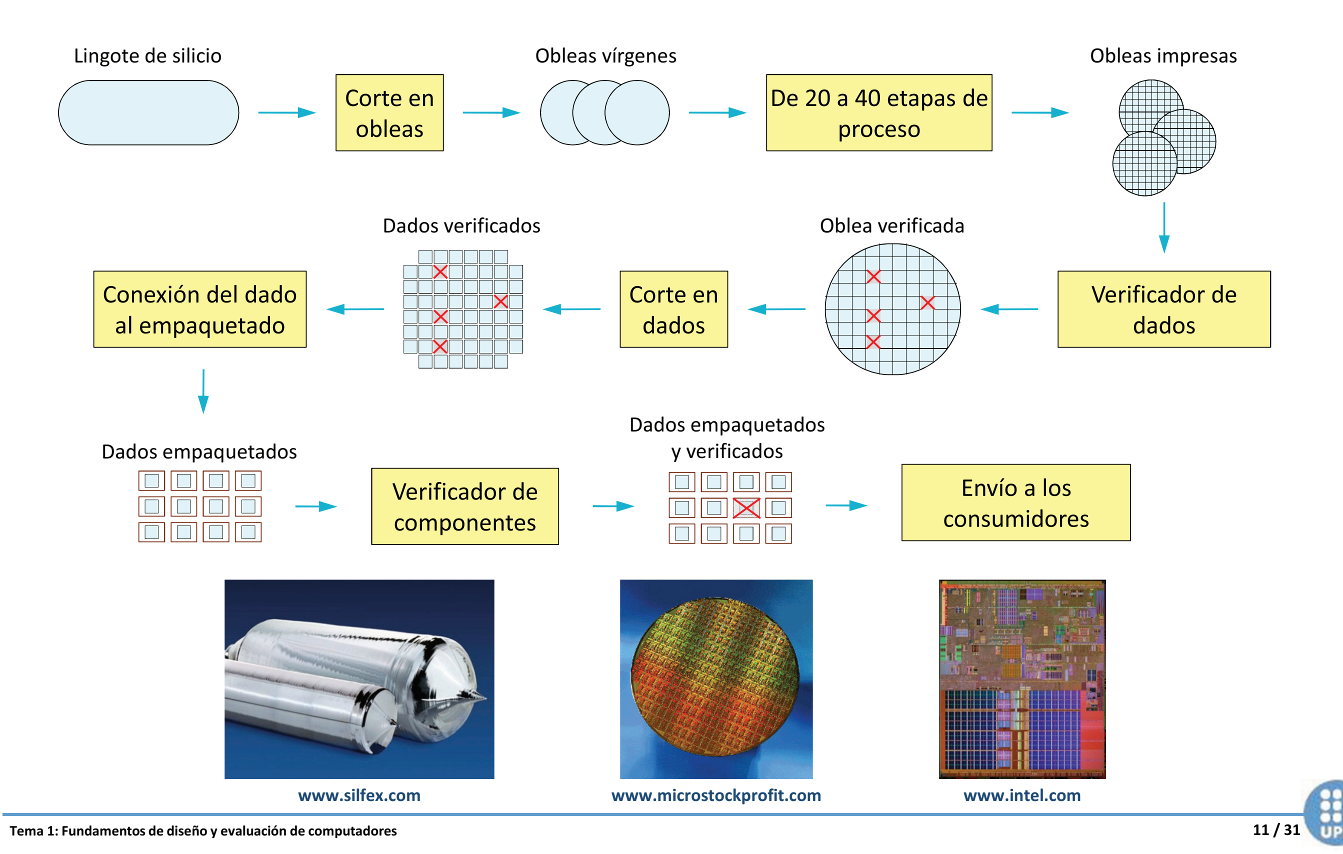
- Tecnología utilizada: **CMOS**
- Elemento básico: **el transistor**



- La tecnología se identifica por la longitud de la puerta del transistor medida en micrómetros (μm, 10^{-6} m) o nanómetros (nm, 10^{-9} m).

	Tecnología	Ejemplo
1971	10 μm	Intel 4004
1975	3 μm	Intel 8088
1982	1,5 μm	Intel 286
1985	1 μm	Intel 386
1989	0.8 μm	Intel 486
1994	0.6 μm	Power PC 601
1995	0.35 μm	AMD K5
1998	0.25 μm	Alpha 21264
1999	180 nm	Pentium III
2000	130 nm	AMD Athlon XP
2002	90 nm	Pentium 4
2006	65 nm	IBM Cell (PS3)
2008	45 nm	IBM POWER 7
2010	32 nm	Intel i7
2011	22 nm	-
2013	16 nm	-
2015	11 nm	-

Proceso de creación de un chip



Evaluación del coste

- **Factor de yield: fracción de circuitos correctos**
- **Coste de un circuito integrado**

$$\text{Coste de un circuito integrado} = \frac{\text{Coste del die} + \text{Coste de testeo} + \text{Coste de empaquetado y test final}}{\text{Yield final}}$$

- **Coste del die (dado)**

$$\text{Coste del die} = \frac{\text{Coste del waffer}}{\text{Dies per waffer} \times \text{Die yield}}$$

- **Dies per waffer (oblea)**

Waffer area / die area

$$\text{Dies per waffer} = \frac{\text{Area útil}}{\text{Die area}} = \frac{\pi \times (\text{diametro}/2)^2}{\text{Die area}} - \frac{\pi \times \text{diameter}}{\sqrt{2 \times \text{Die area}}}$$

Compensación por los dies incompletos de los bordes

- **Die yield**

$$\text{Die yield} = \text{Waffer yield} \times \left(1 + \frac{\text{defectos por unidad de area} \times \text{die area}}{\alpha}\right)^{-\alpha}$$

  - α = medida de la complejidad, se aproxima al número de máscaras críticas

Latencia y Ancho de Banda

- **LATENCIA**: tiempo que transcurre entre la solicitud de un dato (a memoria por ejemplo) y la disponibilidad del mismo. Se mide en ciclos o unidades de tiempo (s).
- **ANCHO de BANDA**: número de bytes transmitidos por unidad de tiempo. Se mide en KB/s, MB/s, GB/s (siempre potencias de 10).

	Latencia	Ancho de Banda
Memoria DDR3-1600	8,75 ns (10^{-9} s)	12,8 GB/s
Gigabit Ethernet	190µs (10^{-6} s)	1Gb/s
Disco Duro SATA -600	7 ms (10^{-3} s)	145 MB/s

Latencia y Ancho de Banda

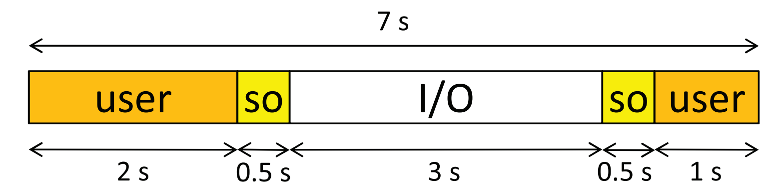
Barcelona 300 Km Zaragoza

LATENCIA
- Ping a www.unizar.es: **10 ms**
- Ferrari F1 (a 300 Km/h): **3.600 s**
- Camión Volvo FH16 (a 100 Km/h): **10.800 s**

ANCHO de BANDA
- ADSL a 20 Mb/s: **2,5 MB/s**
- Ferrari F1 (transportando 1 HD de 1 TB): **277,8 MB/s**
- Camión Volvo FH16 (transportando 34.000 HD de 1 TB): **3,15 TB/s**

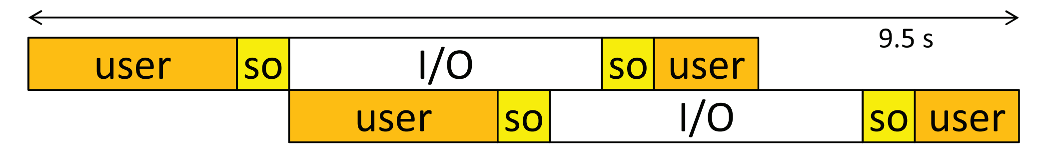
Productividad vs Tiempo de respuesta

- **Tiempo de respuesta (wall time)**



  - Tiempo de usuario: 3s
  - Tiempo de sistema: 1s
  - Tiempo de CPU: 4s
  - Tiempo de respuesta: 7s
  - Throughput: 1 proceso/7 s = 0.14 procesos/segundo

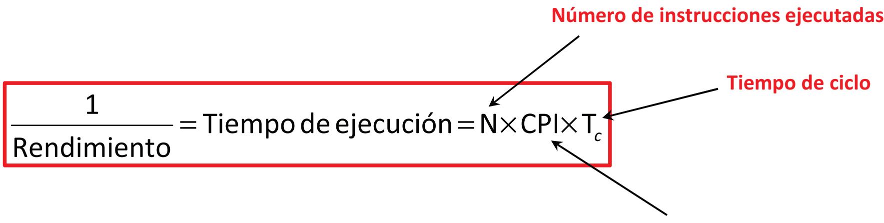
- **Productividad (throughput) = trabajo/tiempo**



  - Throughput = 2 procesos/9.5 s = 0.21 procesos/segundo

Métricas de Rendimiento

- **Rendimiento de un procesador**



$$\frac{1}{\text{Rendimiento}} = \text{Tiempo de ejecución} = N \times CPI \times T_c$$

↑Tiempo ejecución ⇒ ↓Rendimiento

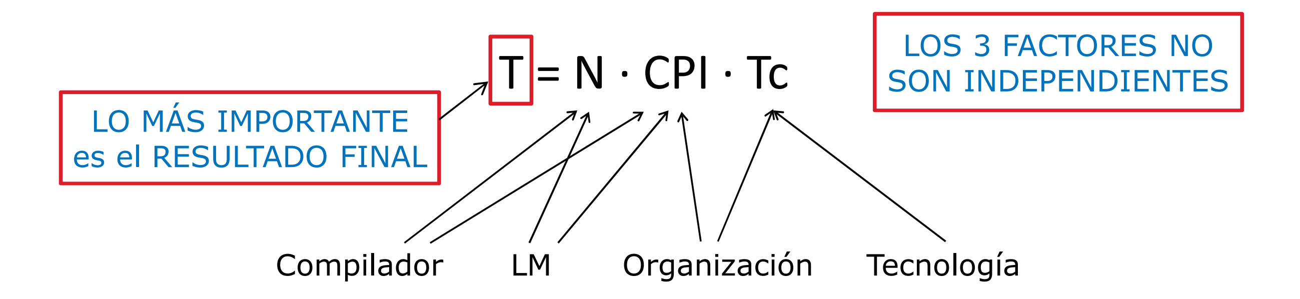


$$\frac{1}{\text{Rendimiento}} = \frac{\text{tiempo}}{\text{Programa}} = \frac{\text{instrucciones}}{\text{Programa}} \times \frac{\text{ciclos}}{\text{instrucción}} \times \frac{\text{tiempo}}{\text{ciclo}}$$

Métricas de Rendimiento

Rendimiento de un procesador

- Reducir el tiempo de ejecución ⇒ Se puede actuar en cualquiera de los 3 factores



Otras métricas de Rendimiento

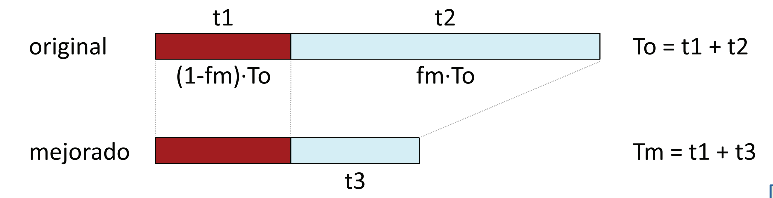
- MIPS: Millones de instrucciones por segundo
- MFLOPS: Millones de operaciones en punto flotante por segundo

Métricas de Rendimiento

Ley de Amdahl

- Expresión formal del sentido común.
- Si aplicamos una mejora, la ganancia obtenida depende de la fracción del tiempo original donde se usa esa mejora.

$$\text{Ganancia} = \frac{T_O}{T_M} = \frac{\text{Tiempo Original}}{\text{Tiempo Mejora}} = \frac{\text{Rendimiento Mejora}}{\text{Rendimiento Original}}$$



fm, fracción mejorada
gm, ganancia de la fracción mejorada

$$\text{Ganancia} = \frac{T_O}{T_M} = \frac{T_o}{(1-fm)\cdot T_O + t3} = \frac{1}{(1-fm) + fm/gm}$$

$$gm = \frac{t2}{t3} = \frac{fm \cdot T0}{t3} \Rightarrow t3 = \frac{fm \cdot To}{gm}$$

Métricas de Rendimiento

Ley de Amdahl

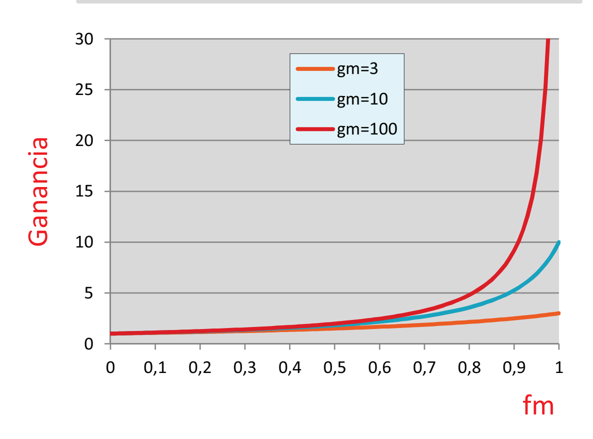
$$\text{Ganancia} = \frac{1}{(1-\text{fm}) + \text{fm}/\text{gm}}$$

Para que la ganancia obtenida sea significativa (cercana a gm) fm ha de ser prácticamente 1.

- Calcular fm para que la ganancia sea gm/2

$$\text{Ganancia} = \frac{1}{(1-\text{fm}) + \text{fm}/\text{gm}} = \frac{\text{gm}}{2} \Leftrightarrow \text{fm} = \frac{\text{gm}-2}{\text{gm}-1}$$

gm	5	10	50	100	1000
fm	0,75	0,88	0,979	0,984	0,9989
Ganancia	2,5	5	25	50	500



Métricas de Rendimiento

Ley de Amdahl

- Regla de diseño: Optimizar el caso frecuente.
- Ejercicio, **¿Qué ocurre cuando gm → ∞?**

$$\Rightarrow \text{Ganancia} = \frac{1}{(1-\text{fm})}$$

fm	0,2	0,5	0,75	0,9	0,99
Ganancia	1,25	2	4	10	100

- Pero no hay que mejorar excesivamente el caso común, incrementos modestos de fm son más efectivos que grandes mejoras en gm.
- La ley de Amdahl se puede aplicar a múltiples situaciones.

Comparación de Rendimientos

- **Para comparar el rendimiento de 2 computadores usaremos el tiempo de ejecución:**

$$\text{Ganancia (Speedup)} = \frac{T_A}{T_B} \quad \begin{array}{l} >1 \Rightarrow \textbf{B es más rápido que A} \\ <1 \Rightarrow \textbf{B es más lento que A} \end{array}$$

  - Si un programa P tarda 4,5 segundos en el computador A y 2 segundos en el computador B:

$$\text{Ganancia} = \frac{T_A}{T_B} = \frac{4{,}5}{2} = 2{,}25 \quad \Rightarrow \textbf{B es 2,25 veces más rápido que A, usaremos 2,25x}$$

  - También podemos usar porcentajes:

$$\left(\frac{T_A}{T_B} - 1\right) \cdot 100$$

    ✓ Una ganancia del 125% ⇒ B es 2,25 veces más rápido.



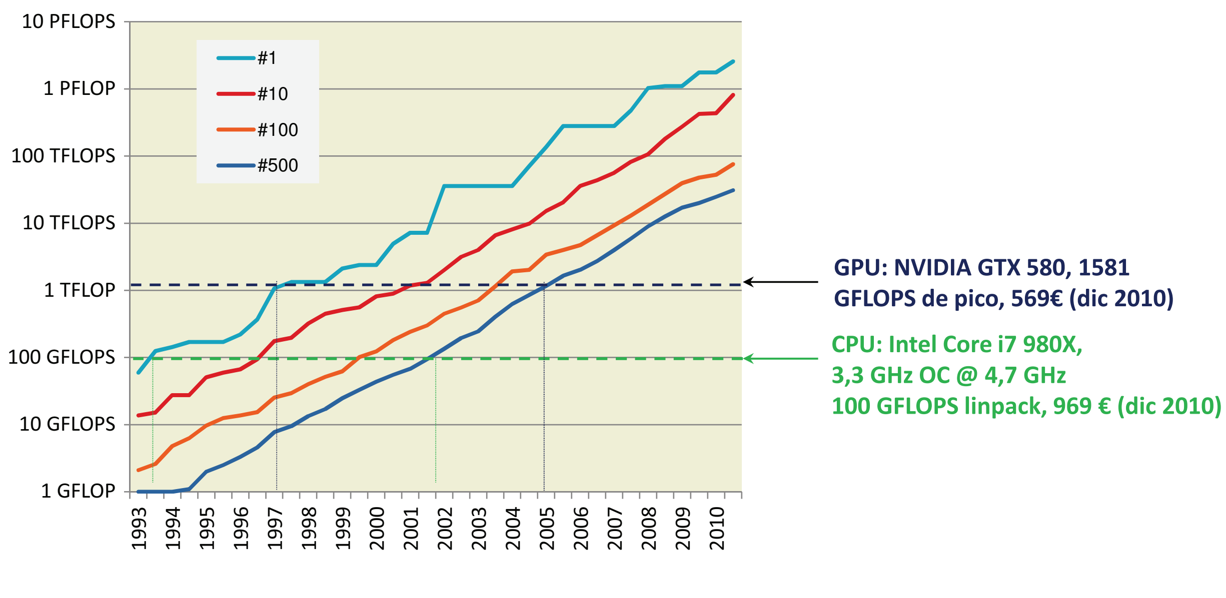
Top500

Rank	Site	Manuf.	Country	Cores	GFLOPS	CPU Family	System Model	OS	Área
1	Oak Ridge National Laboratory	Cray	USA	224.162	1.759.000	AMD x86_64	Cray XT5-HE	Linux	-
5	National SuperComputer Center in Tianjin/NUDT	NUDT	China	71.680	563.100	Intel EM64T	NUDT TH-1 Cluster	Linux	Research
10	Sandia National Laboratories	Sun	USA	41.616	423.900	Intel EM64T	Sun Blade x6275	CentOS	Research
23	Government	Cray	USA	20.960	165.600	AMD x86_64	Cray XT5 QuadCore	CNL	Defense
36	JAXA	Fujitsu	Japan	12.032	110.600	Sparc	Fujitsu FX1	Open Solaris	Aerospace
53	Cray Inc.	Cray	USA	11.076	91.080	AMD x86_64	Cray XT5 SixCore/QuadCore	Linux	Hardware
77	Barcelona Supercomputing Center	IBM	Spain	10.240	63.830	Power	BladeCenter JS21 CLuster	SuSE Linux	Research
78	IBM Poughkeepsie Benchmarking Center	IBM	USA	7.200	63.250	Power	BladeCenter QS22 Cluster	Linux	Software
89	United Kingdom Meteorological Office	IBM	UK	3.520	51.863	Power	pSeries 575	AIX	Weather Forecasting
100	IBM - Rochester	IBM	USA	16.384	47.725	Power	BlueGene/P	CNK/SLES 9	Benchmarking
143	Engineering Company	HP	Germany	5.120	37.153	AMD x86_64	Cluster Platform 4000 BL465c	Linux	Automotive
154	Entertainment Company	HP	USA	5.632	35.936	Intel EM64T	Cluster Platform 3000 BL 2x220	Linux	Entertainment
185	Banking	IBM	Netherlands	5.592	32.278	Intel EM64T	xSeries x3650 Cluster Xeon QuadCore	Linux	Database
193	WETA Digital	HP	New Zealand	5.936	31.527	Intel EM64T	Cluster Platform 3000 BL 2x220	Linux	Media
311	Centre for High Performance Computing	SUN	South Africa	2.624	25.440	Intel EM64T	Sun Blade x6275/X6250	SUSE Linux	Research
330	Digital Media (J)	IBM	USA	8.064	24.670	Intel EM64T	BladeCenter HS21 Cluster, HT Xeon QuadCore	Linux	Gaming
500	University of Reading	IBM	UK	2.800	20.051	Power	BladeCenter JS21 CLuster	Linux	-

- Ranking de los 500 supercomputadores más potentes del mundo. **www.top500.org**
- La lista se actualiza 2 veces al año: junio (ICS) y noviembre (SC)
- LINPACK es la aplicación utilizada para hacer este ranking.

Datos de noviembre de 2009

Top 500



¡ Escala logarítmica !

Consumo

- **Potencia y Energía**
  - Potencia es el trabajo realizado por unidad de tiempo
  - Energía = Integral de la potencia en el tiempo



Potencia es el nivel de consumo

Energía es el área (nivel·tiempo)

  - La Energía se mide en unidades de trabajo: **julios**
  - Potencia es la energía consumida por unidad de tiempo: **watios** (julios/seg)
  - Si la potencia es constante: Energía = Potencia x t
  - La Potencia es importante por razones de disipación térmica
  - La Energía consumida es importante por el coste (económico y/o ambiental) y/o para incrementar la duración de la batería que alimenta al computador.
    - ✓ Batería = Energía

Consumo

- **Energía y potencia eléctricas**

  Potencia = I × V = amperios × voltios = watios

  Energía = P × t = I × V × t = amperios × voltios × segundo = watios × segundo = julios

- **La Potencia consumida por un circuito CMOS tiene 3 componentes:**
  - Conmutación: debido a la conmutación entre niveles de tensión en la carga capacitiva efectiva de todo el chip.
  - Corriente de fugas: los transistores no son ideales.
  - Corriente de cortocircuito: los dos transistores del inversor están activos cuando la entrada cambia de tensión.
- La potencia debida a conmutación es la más importante, aunque la de fugas representa un porcentaje cada vez mayor debido a las reducidas dimensiones de los transistores.



Consumo

- **Potencia y energía de conmutación:**

  $$\text{Potencia} = C \times V^2 \times f$$

  $$\text{Energía} = C \times V^2$$

siendo,

- f, frecuencia
- C, capacidad efectiva equivalente de todo el chip en 1 ciclo (faradios)
- V, tensión de alimentación

- **Potencia de fugas:**

  $$\text{Potencia}_{\text{de fuga}} = I_{\text{de fuga}} \times V$$

Métricas que relacionan Rendimiento y Potencia

- **Métricas de eficiencia**

$$\text{Eficiencia energética} = \frac{\text{rendimiento}}{\text{watio}} = \frac{1}{\text{tiempo} \times \text{watio}} = \frac{1}{\text{Energía consumida}}$$

- Aproximaciones según el tipo de computador
  - ✓ Portátiles (bajo consumo) → 1/Energía = Duración de la batería
  - ✓ Supercomputadores (green 500) → Mflops / W

Green 500

Green 500 Rank	Site	Manuf.	Computer	Country	Mflops/Watt	Power (kW)	Top 500 Rank
1	Forschungszentrum Juelich (FZJ)	IBM	QPACE SFB TR Cluster, PowerXCell 8i, 3.2 GHz, 3D-Torus	Germany	722,98	59,49	110
9	IBM - Rochester	IBM	Blue Gene/P Solution	USA	378,77	126	100
15	IDRIS	IBM	Blue Gene/P Solution	France	378,76	315	32
44	Oak Ridge National Laboratory	Cary	Cray XT5-HE Opteron Six Core 2.6 GHz	USA	253,07	6950,6	1
99	Telecommunications	IBM	BladeCenter HS22 Cluster, Xeon QC GT 2.53 GHz, GigEthernet	Italy	182,58	127,22	402
99	Logistics	IBM	BladeCenter HS22 Cluster, Xeon QC GT 2.53 GHz, GigEthernet	Germany	182,58	127,98	397
99	Data Center (A)	IBM	BladeCenter HS22 Cluster, Xeon QC GT 2.53 GHz, GigEthernet	Hong Kong	182,58	177,05	184
120	Sandia National Laboratories / National Renewable Energy Laboratory	Sun	Sun Blade x6275, Xeon X55xx 2.93 Ghz, Infiniband	USA	173,38	2444,94	10
234	Barcelona Supercomputing Center	IBM	BladeCenter JS21 Cluster, PPC 970, 2.3 GHz, Myrinet	Spain	93,37	683,59	77
235	University of Reading	IBM	BladeCenter JS21 Cluster, PPC 970, 2.5 GHz, Myrinet	UK	93,00	215,6	500
500	IT Service Provider (C)	HP	Cluster Platform 4000 BL465c, Opteron DC 2.4GHz, GigEthernet	USA	13,03	1613,82	479

www.green500.org

Datos de noviembre de 2009

- Ranking de los 500 supercomputadores más potentes del mundo ordenados según eficiencia energética (MFLOPS/watio).
- El supercomputador que más consume (top2 nov 2010): 6950 kW.
- Potencia contratada en un hogar estándar: 5,5 kW
- Potencia de ASCO 2: 1027,2 MW (los 500 supercomputadores del top500 consumen todos juntos alrededor de 300 MW, nov 2010)

Fiabilidad

- **Métricas para caracterizar la fiabilidad:**
  - Fiabilidad: tiempo de funcionamiento continuo sin fallos
    - ✓ MTTF = Mean Time To Failure
  - Tasa de fallos (Failure rate)

$$\lambda = \frac{1}{\text{MTTF}}$$

  - Interrupción del servicio se mide como el tiempo medio necesario para restablecerlo
    - ✓ MTTR = Mean Time To Repair
  - Tiempo medio entre fallos (Mean Time Between Failures)
    - ✓ MTBF = MTTF+MTTR
  - Disponibilidad (availability): Fraccción del tiempo en que un sistema está funcionando.

$$\text{Availability} = \frac{\text{MTTF}}{\text{MTTF}+\text{MTTR}}$$



Fiabilidad

- **El tiempo entre fallos se aproxima a una distribución exponencial donde:**
  - p = probabilidad de que se produzca un fallo
  - λ = 1/MTTF (failure rate)
  - x = tiempo transcurrido

$$p = 1 - e^{-\lambda x}$$

- **¿Cómo calcular el MTTF de un sistema, dado el MTTF de los componentes?**
  - Dados dos componentes con fallos independientes y distribución exponencial
    - ✓ Probabilidades de fallo p_1 y p_2
    - ✓ Tasas de fallo λ_1 y λ_2
    - ✓ Tiempos medios entre fallos $MTTF_1$ y $MTTF_2$
  - probabilidad de que se produzca un fallo es 1 – “probabilidad que no falle ninguno”, o sea:

$$p = 1 - (1 - p_1) \times (1 - p_2) = 1 - e^{-\lambda_1 x} \times e^{-\lambda_2 x} = 1 - e^{-(\lambda_1 + \lambda_2)x}$$

  - Que sigue una distribución exponencial con $\lambda = \lambda_1 + \lambda_2$ de donde se deduce que :

$$\frac{1}{\text{MTTF}} = \frac{1}{\text{MTTF}_1} + \frac{1}{\text{MTTF}_2}$$

Fiabilidad

- **Ejercicio**

  Sistema formado por:
  - 1 CPU (incluye placa base y memoria) MTTF = 1.000.000 horas
  - 2 Discos MTTF = 500.000 horas
  - 1 Fuente de alimentación MTTF = 200.000 horas

  Calcular el MTTF del sistema:

- **La forma de mejorar la fiabilidad es mediante redundancia.**
  - ✓ En tiempo: Repetir un cálculo para comprobar si es erróneo
  - ✓ En recursos: Disponer de componentes extra que reemplazan al que falla
  - Ver ejemplos páginas 26 y 27 de H&P